



Gu, Y., & Yu, G. (2018). : “”. (*English Language Learning*), 2018(10), 39-44.

Peer reviewed version

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the author accepted manuscript (AAM). Please refer to any applicable terms of use of the publisher.

## **University of Bristol - Explore Bristol Research**

### **General rights**

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/pure/about/ebr-terms>

# 如何研究课堂形成性评估

## — 以 " 中国基础教育外语测评研究基金 " 立项课题为例

顾永琦（新西兰惠灵顿维多利亚大学）

余国兴（英国布里斯托大学）

### 摘要

本文简短介绍课堂形成性评估的基本特征，以及课堂形成性评估研究的原则、研究设计和数据分析。我们以“中国基础教育外语测评研究基金”首期课题（2017）的课堂录像为基础，点评课堂教学过程中出现的测评活动，分析这些课堂评估活动在多大程度上具有形成性。

## 1 绪论

“中国基础教育外语测评研究基金”的成立激励了一线外语老师对课堂评估的研究热情。首批科研项目已经或正在结出累累硕果。与此同时，老师在课题实施过程中也遇到了一些困惑，特别是如何把课题研究和课堂测评有机结合、融为一体，如何从研究的视角挖掘和反思自己的课堂测评活动。本文根据作者在第二届英语教学与测评学术研讨会的问题发言整理（有删减），希望能够帮助一线英语教师更好地了解课堂评估，提高课堂评估的效率，同时也提高教师的测评研究素养，培养教师用研究的精神和视角反思自己的课堂教学与测评活动，最终改善学生的英语学习课堂环境、提高学生的英语水平。

## 2 课堂形成性评估的基本特征

课堂形成性评估有哪些最基本的特征？它有哪些作用？为什么很重要？它与我们通常所说的语言测试有哪些根本不同点？如果我们去听一堂非常普通的、日常的英语课，有哪些课堂活动我们可以定义为形成性评估？首先我们来看一下课堂形成性评估的基本特征。课堂形成性评估通常是由教师主导的、由教师发起或协调的，嵌入在教学活动中并与教学活动有机融合的评估活动。课堂形成性评估与大规模考试最大的区别在于前者不需担负或者提供大规模考试以“问责”为目的的信息；课堂形成性评估以“反馈”以及基于反馈的教学与学习活动作为其根本的机制。反馈能帮助学生认识到自己现有的知识、理解或者技能与

预期目标之间的差距，并且指导学生通过可操作的、必要的行动达到预期的目标。形成性反馈不仅仅是帮助学生意识到自己的知识差距，更重要的是为学生提供积极的、正面的信息帮助他们进步和改善成绩。反馈本身并不具备“形成性”，只有当学生接受了反馈并且内化了反馈，并根据反馈而采取调整措施，反馈才有可能有效地促进教与学的过程（Rea-Dickins & Gardner, 2000）。如果学生不能理解和使用这个反馈，或者没有机会内化这个反馈，不管这个反馈有多么的具体细致，都不能起到提高学生学习成绩的作用（Sadler 1989）。将形成性反馈有机融合到课堂活动中，从而为学生提供一个与老师进一步交流的机会，来展示自己在多大程度上已经掌握了所学的内容，这样的交流机会则会进一步促进学习（Black & Wiliam 1998）。如果学生自己能够产生、展示这样的相关信息，这个过程我们称之为自我监控；但是如果这样的相关信息来源于外部，如老师，这个过程我们称之为反馈（Sadler 1989）。

从老师的视角来看，课堂形成性评估通常包括下列活动：

- 与学生共享学习目标
- 提供形成性反馈
- 活跃的师生以及学生与学生之间的有效互动
- 有效的、有策略的提问和提示
- 在整堂课上不断地调整自己的教学、协调学生的学习
- 通过学生的自我反思（包括自我监控和同伴互评）来促进学生的学习
- 为学生提供各种深入学习的机会

在课堂形成性评估中，老师担当着两种角色。首先老师是语言的协调者和内容的提供者，为学生提供形成性的支持与反馈；再者老师是一个评判员，需要做出判断和决定，判断学生的语言是否正确合适。老师应遵循形成性评估的十条基本原则（Assessment Reform Group, 2002）：

- 1) 形成性评估应该融入教学计划，是教学计划的重要组成部分
- 2) 形成性评估应该聚焦学生怎么学习
- 3) 形成性评估应该是课堂活动的核心
- 4) 形成性评估应该是教师的核心素养和能力
- 5) 形成性评估应该是善意的、建设性的，因为任何评估都会影响情绪
- 6) 形成性评估应该考虑学生的学习热情
- 7) 形成性评估应该促进师生对学习目标和评估标准的认识和认同

- 8) 形成性评估应该给予学生如何改进学习的建设性指导
- 9) 形成性评估应该培养学生自我评估的能力，帮助学生反思和自我管理
- 10) 形成性评估应该认可学有所长所短，顾及所有学生，各个层次的学生

### 3 研究课堂形成性评估

我们在指导首期课题时，经常听到老师问三个问题：我这么做算不算研究？研究到底长啥样？我应该怎么研究课堂形成性评估？首先，提出研究问题是课堂评估研究的最关键的第一步。研究问题在很大程度上决定了研究的类别。研究问题必须是细化的、具体的、可操作的、可实证回答的。其次，相对于语言教学与测评其他领域的研究，我们认为课堂形成性评估研究可用的数据收集和分析方法比较有限。表 1 列出了课堂评估研究的大致分类和相关的数据收集和分析方法。根据研究的初始目的和研究问题，我们可以选用相应的数据收集和分析方法。

表 1：研究分类与相应的数据收集和分析方法

研究目的类别	数据收集	数据分析
描述型、探索型研究	<ul style="list-style-type: none"> <li>• 课堂录音/录像</li> <li>• 课堂观察列表</li> <li>• 师生访谈</li> <li>• 师生日志</li> <li>• 课堂备课计划、练习和其它相关材料</li> <li>• 问卷（如果是规模较大的研究）</li> </ul>	<ul style="list-style-type: none"> <li>• 编码和分析：如 IRF 语篇结构、活动类型、形成性评估的内容和种类</li> <li>• 描述性统计</li> </ul>
相关性研究	<ul style="list-style-type: none"> <li>• 问卷调查</li> <li>• 考试</li> </ul>	<ul style="list-style-type: none"> <li>• 相关性分析</li> <li>• 多元回归</li> </ul>
因果关系研究：实验	<ul style="list-style-type: none"> <li>• 干预、前测、后测，与非实验组的对比</li> </ul>	<ul style="list-style-type: none"> <li>• T-test, ANOVA, ANCOVA, MANOVA</li> </ul>
因果关系研究：行动研究	<ul style="list-style-type: none"> <li>• 发现问题 -&gt; 实施创新方案 -&gt; 观察 -&gt; 评估 -&gt; 实施改进方案 -&gt; 观察 -&gt; 评估 -&gt; 实施改进方案 -&gt; 观察 -&gt; 评估 -&gt;</li> </ul>	<ul style="list-style-type: none"> <li>• 详细描述研究过程</li> <li>• 统计对比（如有可能）</li> </ul>

对比研究	<ul style="list-style-type: none"> <li>• 所有可比对的数据</li> </ul>	<ul style="list-style-type: none"> <li>• 质性和量化对比（如不同性别、成绩水平、农村城市之间的对比）</li> </ul>
------	--	---

我们分析了 " 中国基础教育外语测评研究基金 " 首期立项的八个课题的研究问题，发现其中有些问题其实并不能称为研究问题。我们把这些问题归为三类 " 教学方法问题 "，" 文献综述问题 "，和 " 其它拓展问题 "（如研究结果的应用）。研究问题与这三类问题最本质的区别在于研究问题是可以通过数据来回答的。

我们以一个立项课题中的 " 研究问题 " 为例：

如何使诊断结果更好地服务于阅读教学？

这个 " 研究问题 " 本质上来说是一个关于怎么教阅读的问题，根据项目的设计书我们认为它并不是真正意义上的研究问题。我们对这个问题进行重新设计使之成为可回答的研究问题，表 2 列出了我们认为可能的新的研究问题。

表 2：新的研究问题、数据收集和分析方法 (a)

新的研究问题	数据收集	分析
诊断结果是什么？	诊断考试的成绩	<ul style="list-style-type: none"> <li>• 描述性统计</li> <li>• 质性解读、分析诊断考试的结果</li> </ul>
基于诊断结果的阅读教学能够提高学生阅读成绩吗？	<ul style="list-style-type: none"> <li>• 干预方法（如根据诊断结果提供反馈，反馈 + 跟进训练）</li> <li>• 阅读理解考试（前测 + 后测）</li> <li>• 学生访谈</li> </ul>	<ul style="list-style-type: none"> <li>• 统计比较前测与后测之间的成绩差异</li> <li>• 统计比较不同干预的结果</li> <li>• 行动研究和详细描述质性数据</li> </ul>

我们还发现，这些项目中的 '研究问题' 存在另外一个问题，即研究问题过于笼统广泛，夹杂着太多的变量，如下面这个研究问题：

诊断式测评 " 优诊学 " 的反馈是否有助于增强高中生的英语自主学习、包括提高学习写作微技能（如动词搭配，词汇和语法，篇章衔接等），进而提高英语写作水平？

我们知道这个项目的总体目标以及这个研究问题所涉及的一些变量，可以对这个研究问题进行修改。表 3 列出我们认为更加具体的、可操作的研究问题。

表 3：新的研究问题、数据收集和分析方法 (b)

新的研究问题	数据	分析
诊断式测评 " 优诊学 " 的反馈是否有助于增强高中生的英语自主学习能力？	<ul style="list-style-type: none"> <li>• 优诊学的反馈</li> <li>• 诊断测试前后学生访谈或者自主学习问卷调查</li> </ul>	<ul style="list-style-type: none"> <li>• 行动研究</li> <li>• 对接分析反馈内容和模式与自主学习行为之间的关系</li> </ul>
诊断式测评 " 优诊学 " 的反馈是否有助于提高高中生的英语写作水平？	<ul style="list-style-type: none"> <li>• 优诊学的反馈</li> <li>• 诊断测试前后学生写作能力的测量或者学生写作过程档案袋材料</li> </ul>	<ul style="list-style-type: none"> <li>• 行动研究</li> <li>• 对接分析反馈内容和模式与学生语言水平的关系</li> <li>• 详细描述反馈与学生写作进步之间的关系</li> </ul>
诊断式测评 " 优诊学 " 的反馈是否有助于提高写作微技能（如动词搭配，词汇和语法，篇章衔接等）？	<ul style="list-style-type: none"> <li>• 优诊学关于各微技能的反馈</li> <li>• 各微技能的多次测试</li> </ul>	<ul style="list-style-type: none"> <li>• 行动研究</li> <li>• 跟踪反馈调查与微技能进步之间的关系</li> </ul>
写作微技能（如动词搭配，词汇和语法，篇章衔接等）的改进是如何提高学生的总体英语写作水平的？	<ul style="list-style-type: none"> <li>• 各微技能的多次测试</li> <li>• 诊断测试前后学生写作能力的测量或者学生写作过程档案袋材料</li> </ul>	<ul style="list-style-type: none"> <li>• 各分项微技能与写作总体水平之间的相关性统计分析</li> <li>• 详细描述学生在哪些方面进步了</li> </ul>

#### 4 课堂实录分析

我们可以对课堂形成性评估的整个过程进行研究，也可专注评估过程中的某个/些方面。我们可以分析课堂评估的目的，过程，期待的效果，实际的效果等等（Torrance &

Pryor, 2001 ) ; 另外 , 我们还可以从话语分析的视角 , 解读老师的提问方式和内容 (Heritage & Heritage 2013) 以及开启 - 回答 - 反馈(initiation-response-feedback)模式 (Sinclair & Coulthard 1975)。下面我们以 " 中国基础教育外语测评研究基金 " 的某一项目录制的一堂 40 分钟的课为例来演示如何发现和分析课堂上发生的形成性评估。这是一堂高二年级的通过写作来教学词汇的英语课 , 使用的教材是人教社 New Senior English for China , 教学内容是第六册第三单元的 a letter for advice。经过优诊学首轮诊断测试后老师发现学生在写作时使用词块的能力较弱 , 因此这节课的重点是教学如何使用下面四个词块 : decide on, every time, instead of, if you feel

表 4 : 课堂实录 ( 部分 )

话轮	时间	内容
1	0:06.4 - 0:15.2	S: Stand up.
2	0:15.1 - 0:18.0	T: Good morning, everyone.
3	0:18.0 - 0:19.8	SS: Good morning, professor.
4	0:19.8 - 0:34.9	T: (inaudible) As you know, everyone wants to live a healthy life. But, why? because without a healthy body, nothing is impossible. Yes?
5	0:34.8 - 0:36.2	SS: Yes.
6	0:36.2 - 0:52.7	T: But how? Can you give me some suggestions? There is no need to put up your hand, just stand up.
7	0:52.7 - 1:09.2	S: Ok. first you need to have a balanced diet; and live a regular life.
8	1:09.2 - 1:10.8	T: Next step?
9	1:10.8 - 1:13.6	S: You should have a positive attitude towards life.
10	1:13.6 - 1:17.2	T: Yes.
11	1:17.2 - 1:24.5	T: You are the best. Anybody else?
12	1:24.5 - 1:47.2	S: We need to sometimes (inaudible). enjoy the happiness and challenge in our everyday life.
13	1:47.2 - 1:51.5	T: That's right. next one?
14	1:51.4 - 2:03.5	S: Exercise
15	2:03.4 - 2:18.6	T: Do exercise. Very good. Now we can say you'd better be positive, right? and (...) good habit, or do exercise, work out every day. Now my question is: do you take exercise every day?

我们提出四个研究问题 :

- 1) 这节课中有哪些评估行为 ?
- 2) 这节课中老师是通过什么方法知道学生的理解和学习情况的 ?
- 3) 这节课中老师使用了哪些反馈 ?

#### 4) 这节课中在反馈之后老师采取了哪些跟进措施？

### 研究结果

在这堂课中老师的评估行为有事先计划好的、也有事先未计划好的，但绝大多数是事先未计划好的课堂评估。表 5 列出了这节课中发生的一些评估行为，以及各评估行为的步骤。事先计划好的评估行为主要发生在这堂课快要结束的时候（从第 141 个话轮开始）。老师介绍评分标准，要求学生自评和同学互评，之后分享“最佳写作”。

表 5 课堂上发生的评估行为

本课结构				课堂评估		
话轮	时间	课堂活动	互动模式	引导出理解程度或学习状况	反馈	跟进
4-18	2 min	热身: 由学生针对如何过健康生活提建议, 引出主题: 提建议	T→SS, T→S, SS→T, S→T	提问	确认: 'yes', 'ok', 'that's right' 表扬: 'you are the best'.	无
19-23	1 min	读李华的求救信	T→SS, T→S, SS→T, S→T	<ul style="list-style-type: none"> <li>朗读</li> <li>提问 ('Is it clear?')</li> </ul>	表扬: 'Your tone and pronunciation is amazing'	无
23-54	5 min	给李华写回信。准备 1: 在课文 19 页找出 do's and don'ts	T→SS, T→S, SS→T, S→T	在课文中找出句子并读出来	确认: 'yes', 重复学生原话 表扬: 'good', 'wonderful' 提高认识: 把目标词块写在黑板上; 'These are very useful suggestions; and when you give suggestions, these can make your writing very coherent, very smooth like a (...)'.	延后: 在以后的练习中运用



55-78	3 min	给李华写信。准备2: 填空, 提供半句话, 需要填充的半句翻译空格后面的中文。	T→SS, T→S, SS→T, S→T	<ul style="list-style-type: none"> <li>根据中文或图片填空</li> <li>提问 ('Understand me?')</li> <li>利用手势让学生回答</li> </ul>	<b>确认:</b> 'yes', 重复学生原话 <b>表扬:</b> 'good', 'very good' <b>提高认识:</b> 'you should remember 'feel like doing something' not 'instead of' or 'want to'.' 	延后: 在以后的练习中运用
79-106	3 min	给李华写信。准备3: 造句以图片代替中文	T→SS T→S S→T SS→T	<ul style="list-style-type: none"> <li>根据图片提示, 运用所学词块提建议</li> <li>提问 ('yes?')</li> <li>请求 ('once again')</li> <li>用升调示意其它答案 ('get up and ...?')</li> </ul>	<b>确认:</b> 'yes', 重复学生原话 <b>表扬:</b> 'good', 'perfect' <b>提高认识:</b> 'I think another word is very useful 'remind" 	延后: 在以后的练习中运用
107-127	5 min	给李华写信。准备4: 小组讨论为李华提建议	S→S, T→S, S→SS, T→SS	<ul style="list-style-type: none"> <li>小组讨论</li> <li>课堂观察</li> <li>小组与全班分享</li> </ul>	<b>表扬:</b> 竖大拇指, 'very good idea' <b>提高认识:</b> 'You should use as many expressions on the blackboard as possible'; 'If you feel' and 'instead of' is frequently used here, right?' <b>确认:</b> 'yes'; 重复学生原话: 'sleep early. OK' <b>继续探测:</b> 'not to sleep. How?' 'go to bed early. That's the same, right?' 	延后: 在下面的写作中运用
128-140	12 min	给李华写信提供建议	T→SS T→S	<ul style="list-style-type: none"> <li>在工作表上写作</li> <li>课堂观察</li> </ul>	<b>提醒:</b> 'write the answer on the right part. Right, not left' 	运用所学的词块

					注重形式: 'add a comma here'	
141-150	2 min	评估回信 1: 介绍评价标准+老师示范	T→SS SS→T	提问: 'it's used properly, yes?' 'how many points?'		
151-158	2 min	评估回信 2: 个人自评与同学互评	T→SS S→S	T: 课堂观察 SS: 阅读自己的和其他小组成员的回信	任务管理: 'exchange yours with your partner and then choose whose work is the best'	
159-217	7 min	评估回信 3: 全班分享'最佳'回信	T→S T→SS S→SS	S: 朗读 T: 提问 'why do you think his is the best?'	确认: 'yes'; 重复学生原话 改正重说: 'that means most expressions are covered' 转移注意力: 'good handwriting. Now we should assess it in terms of the vocabulary or words' 鼓励: 'don't be nervous'	作业

形成性评估最关键的一步是要知道学生的理解和学习状况，有点像传统的考试或小测验的功能，但是形成性评估的方法则远多于传统的测试。课堂上我们可以使用各种各样的方法来了解学生的理解和学习状况，既可以是精确的也可以是模糊的，既可以是正式的也可以是非正式的；可以是看看学生的眼神，也可以是正式的考试。在这堂课中，老师最多使用的是全班提问和个别同学提问来了解学生的理解和学习状况。课堂任务包括小组讨论，朗读，和个人写作。老师也通过观察来了解学生学习这四个词块的状况。反馈是形成性评估的核心。在这堂课中，老师使用最多的反馈是“确认”和“表扬”。确认通常是用“*Yes*”、“*That's right*”或重复一遍学生所说的内容来达到的；表扬大部分是口头的，如“*good*”、“*perfect*”、“*wonderful*”、“*very good*”，老师偶尔也竖大拇指来表示赞赏。老师反馈的主要目的是要提高学生对这四个词块的认识。确认和表扬尽管可以激励学生，但这些反馈行为本身并不能确保评估行为具有形成性的功能。形成性课堂评估的最后

一环节是 " 跟进实施行动 " 。反馈只能告诉学生和老师需要采取什么行动，真正要使评估具有形成性则必须跟进实施具体的行动。这堂课上跟进行动非常有限，因此我们认为这些课堂评估活动大都不具备形成性。【对这堂课具体的分析结果因期刊篇幅有限不做详细报告】

## 5 结束语

本文简短介绍了课堂形成性评估的最基本特征，指出课堂形成性评估是教学过程的有机组成部分。我们以一堂 40 分钟的课堂录像为例来说明研究问题的提出和课堂数据的采集和分析方法。我们认为课堂形成性评估应该是一系列螺旋式上升的、完整的评估环，每一环都包括了解学生现有的学习成绩和表现，提供反馈，设计跟进活动从而减小现有水平与期待水平的差距。如果缺少了最后一个跟进环节，反馈和评估就很难具有形成性。

尽管我们对课堂录像的分析让我们能更好地了解课堂上发生的评估行为，并有机会给老师提出一些课堂评估的改进意见，但是我们必须承认我们的分析有较大的局限性，因为我们的分析是主观的判断，并且我们也没有和老师核实我们的判断是否正确。老师作为课堂评估的主体、发起者、实施者，可以和科研人员一起设计分析课堂形成性评估来增强他们的评估素养。我们认为 " 中国基础教育外语测评研究基金 " 实行的中学一线教师和大学科研人员合作研究的模式可以为一线教师提供一个职业发展的平台(Cochran-Smith & Lytle, 1999; Stremmel, 2007)，同时也保证研究项目的成功开展和研究结果的有效阐述，是一个双赢机制。

## 引用文献

- Assessment Reform Group (2002). Assessment for Learning: 10 principles.
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning, *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74, DOI: 10.1080/0969595980050102
- Cochran-Smith, M., & Lytle, S. L. (1999). The teacher research movement: A decade later. *Educational Researcher*, 28(7), 15-25. <https://doi.org/10.3102/0013189X028007015>
- Cowie, B., & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education: Principles, Policy & Practice*, 6(1), 101-116. <https://doi.org/10.1080/09695949993026>

- Heritage, M., & Heritage, J. (2013). Teacher questioning: The epicenter of instruction and assessment. *Applied Measurement in Education*, 26(3), 176–190.  
<https://doi.org/10.1080/08957347.2013.793190>
- Rea-Dickins, P., & Gardner, S. (2000). Snares and silver bullets: disentangling the construct of formative assessment. *Language Testing*, 17(2), 215–243.  
<https://doi.org/10.1177%2F026553220001700206>
- Sadler, R. (1998). Formative Assessment: revisiting the territory, *Assessment in Education: Principles, Policy & Practice*, 5 (1), 77–84, DOI: 10.1080/0969595980050104
- Sinclair, J. M., & Coulthard, M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford, UK: Oxford University Press.
- Stremmel, A. J. (2007). The value of teacher research: Nurturing professional and personal growth through inquiry. *Voices of Practitioners*, 2(3), 1–9.
- Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal*, 27(5), 615–631.
- William, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 18–40). New York: Routledge.